



AWSのストリームサービスでつくる、 データ価値を100倍にするデータ基盤

Amazon Web Services Japan, K.K.

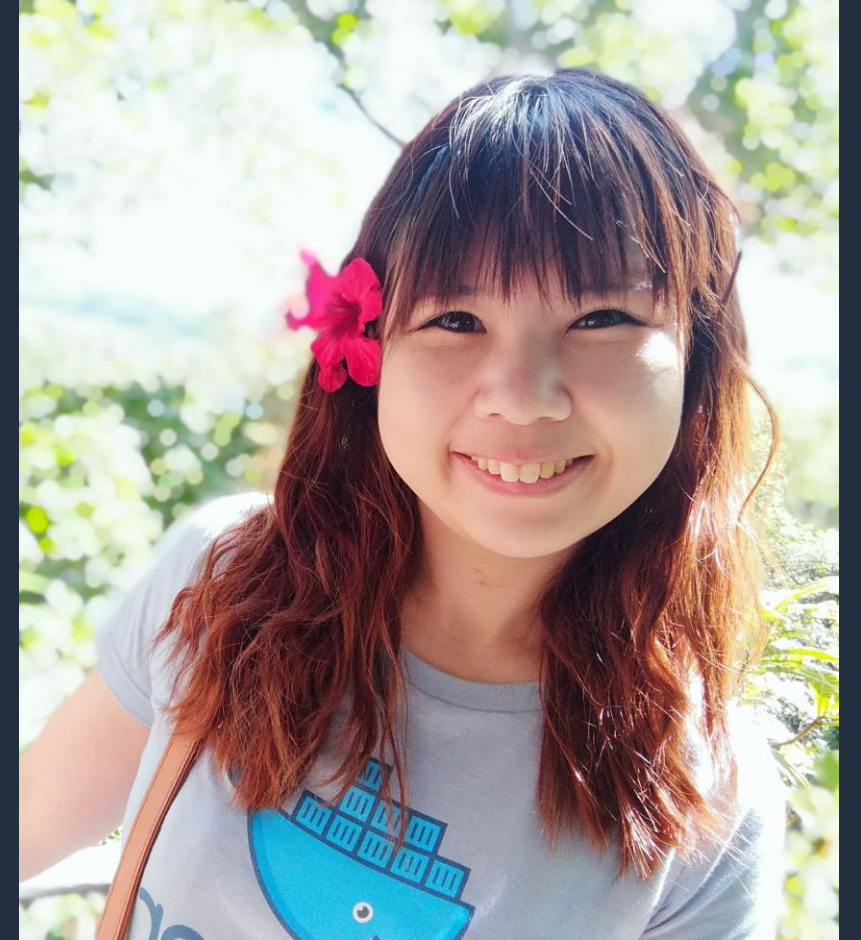
Chie Hayashida

2021/06/24

自己紹介

データアナリティクススペシャリストSA

林田 千瑛(はやしだ ちえ)



なぜストリーム処理が必要なのか

バッチ処理で得られたデータでできること

数分から1日前以前のデータを利用して分析・評価を行う

例)

- 日次・月次でのメトリクスの集計レポート
- 長いスパンでのトレンド分析
- 広告サーバのキャパシティ増減の確認

ストリーム処理で得られたデータでできること

数ミリから数秒単位のレイテンシでデータを受け取って処理をする

例)

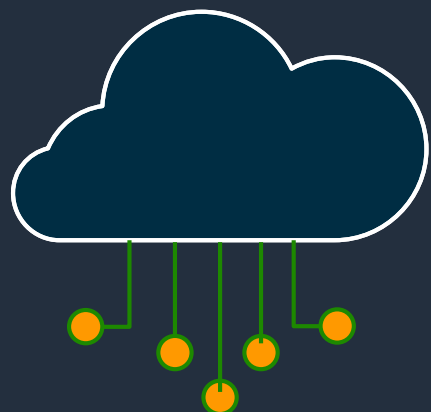
- 偉大な情報を数ミリ秒から数秒のレイテンシでアプリケーションに反映
- 数秒前のデータをダッシュボードに反映
- インシデント時に即座の原因分析

ストリームデータを用いたログ分析の例



ビジネス分析

- ユーザの動向をリアルタイムに確認して施策に反映
- 商品の売上を監視して在庫が無くなりそうになったらアラート
- 金融商品の取引状況を監視
- マシン設備の稼働状況を監視



アプリケーション/インフラログの分析

- レイテンシやエラーレートの急増を検知
- アプリケーションエラーをリアルタイムにドリルダウン分析
- セキュリティインシデント監視/異常検知



セキュリティ

- 疑わしいユーザ動向の検知（不正ユーザの早期発見・原因分析）
- 侵入直後に侵入者のIPでアクセスされたデータを特定

ストリームデータを用いたログ分析の例



ビジネス分析

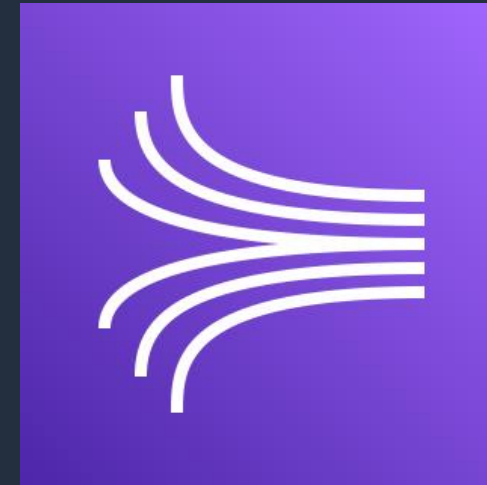
- ユーザの動向をリアルタイムに確認して施策に反映
- 商品の売上を監視して在庫が無くなりそうになったらアラート
- 金融商品の取引状況を監視
- マシン設備の稼働状況を監視

アプリケーション/インフラログの分析

ビジネスにおける意思決定の加速
すばやいサービス改善サイクルの実現

Amazon Kinesis

AWSにおけるストリーミングサービス

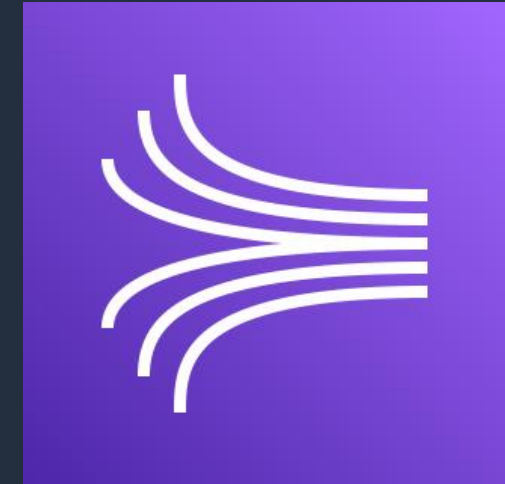


- ストリームの複雑さをクラウド側で吸収
- 価値あるリアルタイムデータを低いコストでビジネス利用

Amazon Kinesis

AWSにおけるストリーミングサービス

- ✓ インフラ管理不要
- ✓ スケーラブル
- ✓ 高い耐障害性
- ✓ セキュア
- ✓ AWSや3rdパーティサービスとの接続性
- ✓ 利用した分のみ課金
- ✓ さまざまなユースケースを網羅する機能



分析対象データの例

IoT & Mobile



- 自動車
- スマートホームデバイス
- IoT搭載ツール
- 製造設備
- モバイルアプリ

IT & DevOps



- データベース
- ロードバランサ
- ネットワークSW
- デプロイツール
- サーバ

Application & Cloud



- アクセスログ
- 環境設定変更ログ
- Webアプリケーションログ
- ビジネスアプリケーション
- コンテナフレームワーク

さまざまなタイプのデータ収集に対応

Producers



AWS SDK

- サーバやモバイルからAPIを通じて直接Publish

AWS Service Integrations

- CloudWatch Logs, AWS IoT, DynamoDB, Amazon Database Migration Service

Kinesis Agent

- ログファイルを監視して新しいレコードを転送

Kinesis Producer Library (KPL)

- ストリームデータを発生させてPublish
- 複数レコードをAggregateできる

3rd party and open source

- Log4j appender
- Apache Kafka
- Flume, fluentd, and more ...

Kinesis Family

どんなユースケースも対応するサービス群

Kinesis Data Streams



ストリームを取得、
加工して転送する機能

Kinesis Data Firehose



ストリームデータを
データストアに
ロードする機能

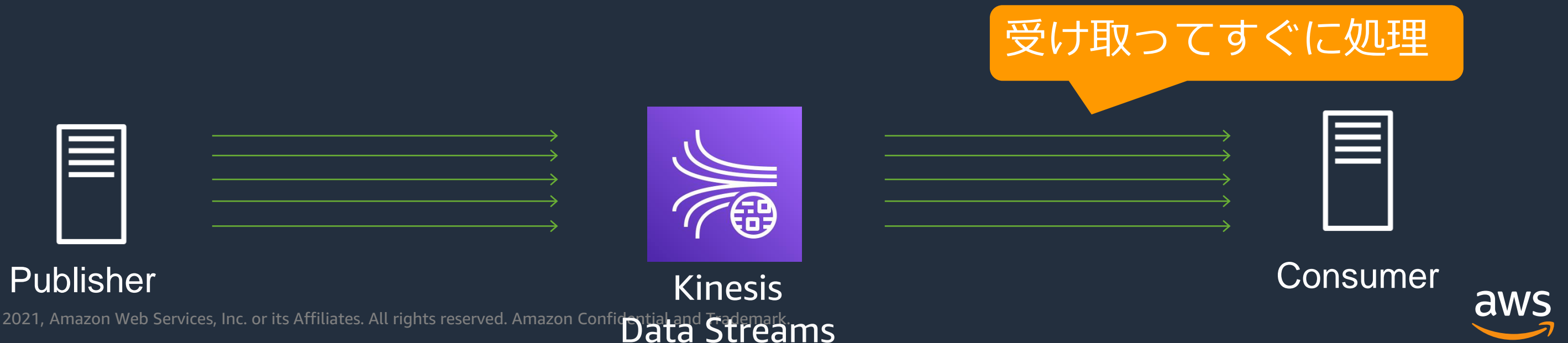
Kinesis Data Analytics



ストリームデータを
分析する機能

Kinesis Data Streams

- データの生成元（Publisher）からストリームデータを受け取り、一時永続化して、データの利用先（Consumer）にデータを渡すサービス
- さまざまなストリーミングサービスと連携
- クライアントライブラリを利用して独自のストリーミングアプリケーション作成が可能
- シャード数の増減によってスループットを調整



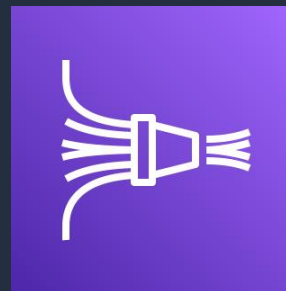
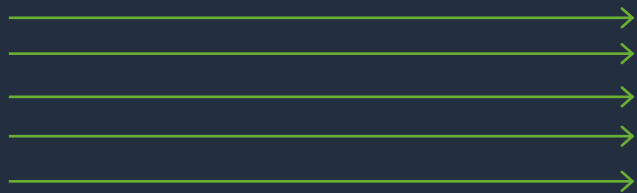
Kinesis Data Firehose

- データ送信元からストリームデータを受け取り、決められたバッファが溜まったタイミングもしくは一定時間ごとにデータストアに格納するサービス
- 格納先のデータストアは、Amazon S3 / Amazon Redshift / Amazon Elasticsearch Service が利用可能
- シャード管理不要

バッファが溜まったタイミング
もしくは決まった頻度で転送



Publisher

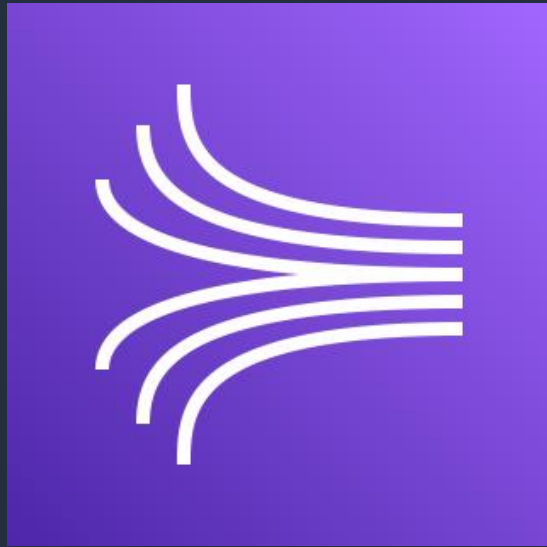


Kinesis
Data Firehose



データストア

ログ可視化からはじめるリアルタイムデータ活用



マネージドストリーム
処理サービス

Amazon Kinesis

マネージド検索サービス

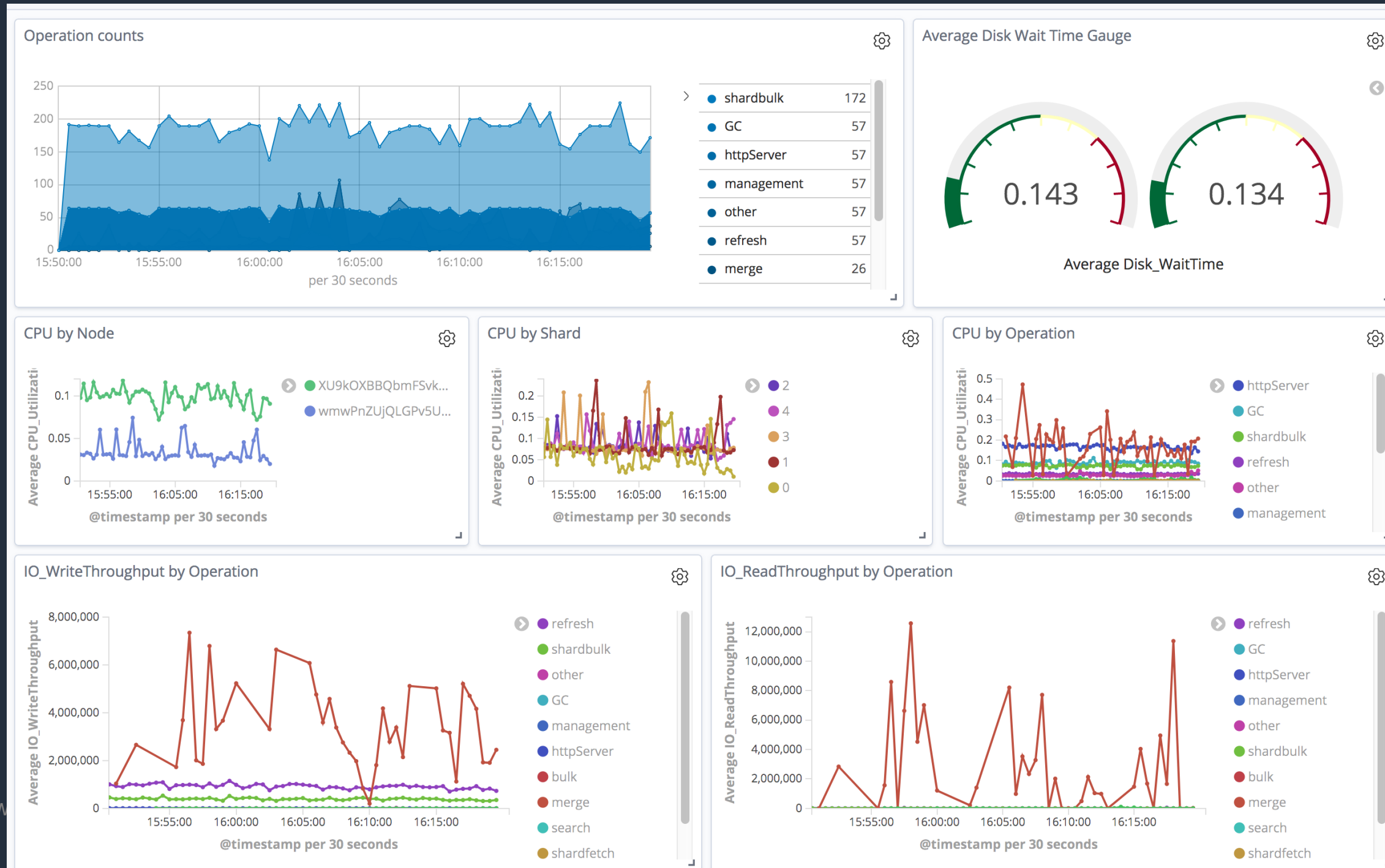
Amazon Elasticsearch
Service

ニアリアルタイムログ分析基盤に Amazon Elasticsearch Serviceが選ばれる理由

- ストリームデータを簡単にInsertして数秒で利用できる
- 可視化機能（Kibana）がビルトイン
- 特定のキーワードに対する集計に向いている
- フルマネージドサービス
- スケーラブル
- 高いコストパフォーマンス

ログ分析ダッシュボード例

サーバリソースの監視

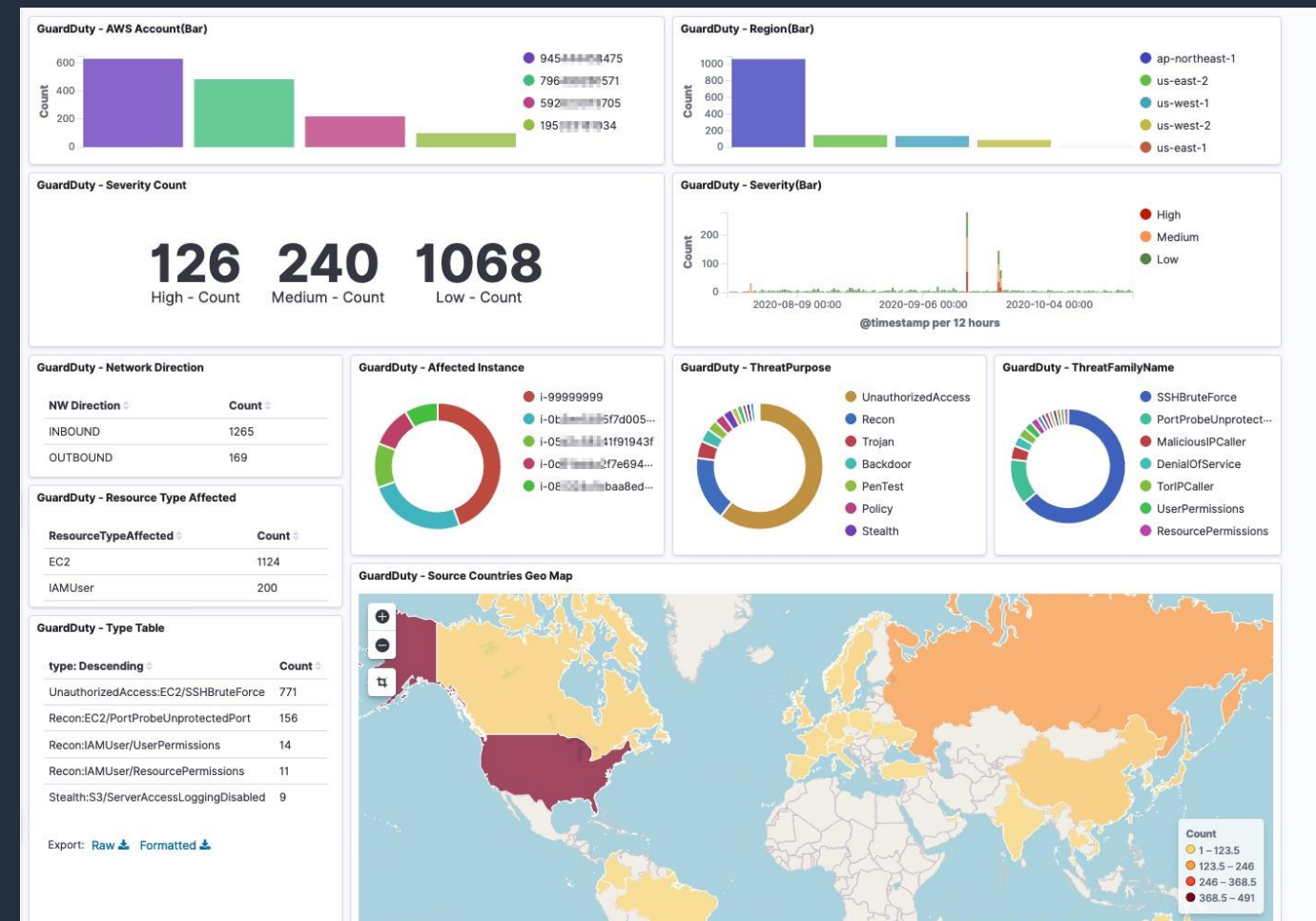


ログ分析ダッシュボード例

セキュリティログの可視化(SIEM)

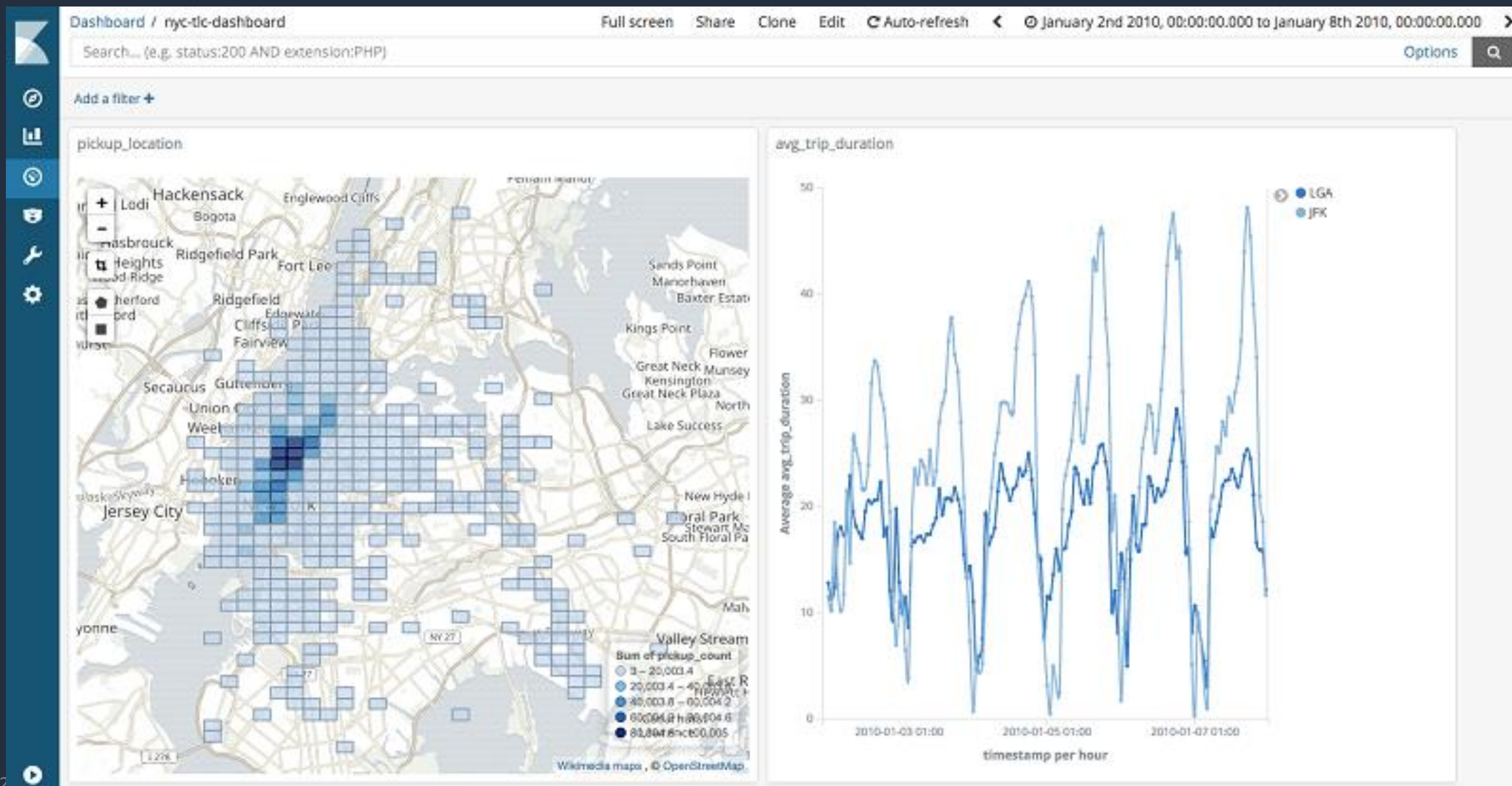


ログをセキュリティ分析の観点から
時系列、地理情報などで視覚化



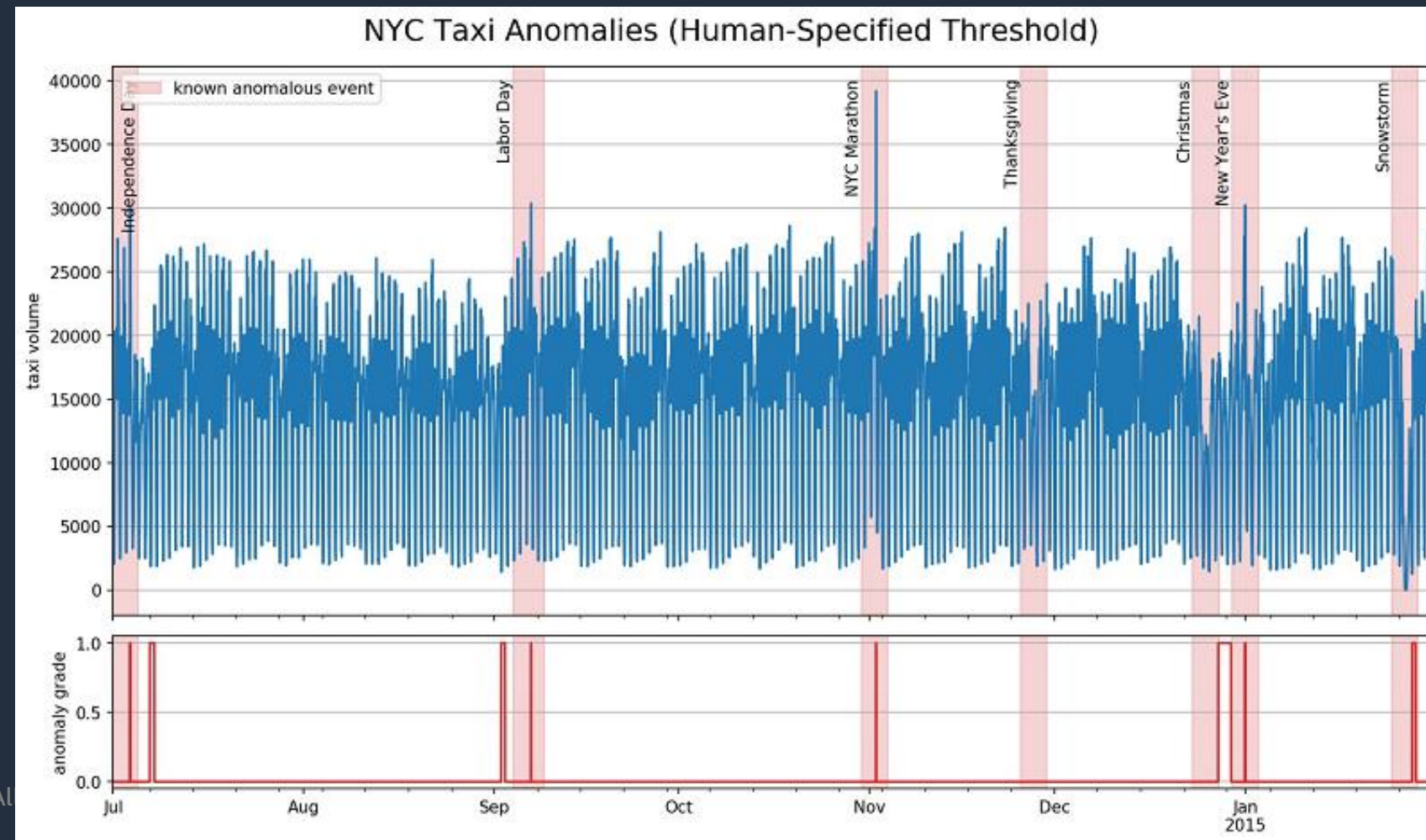
ログ分析ダッシュボード例

アプリケーションログを利用した車両の移動情報の可視化



Anomaly Detection on Amazon Elasticsearch Service

- Random Cut Forest(RCF)を利用して正常の状態を学習してそれと異なる動きをニアリアルタイムで検出
- 検出した異常をアラート機能と連携して通知
- MLの知識なしに利用可能

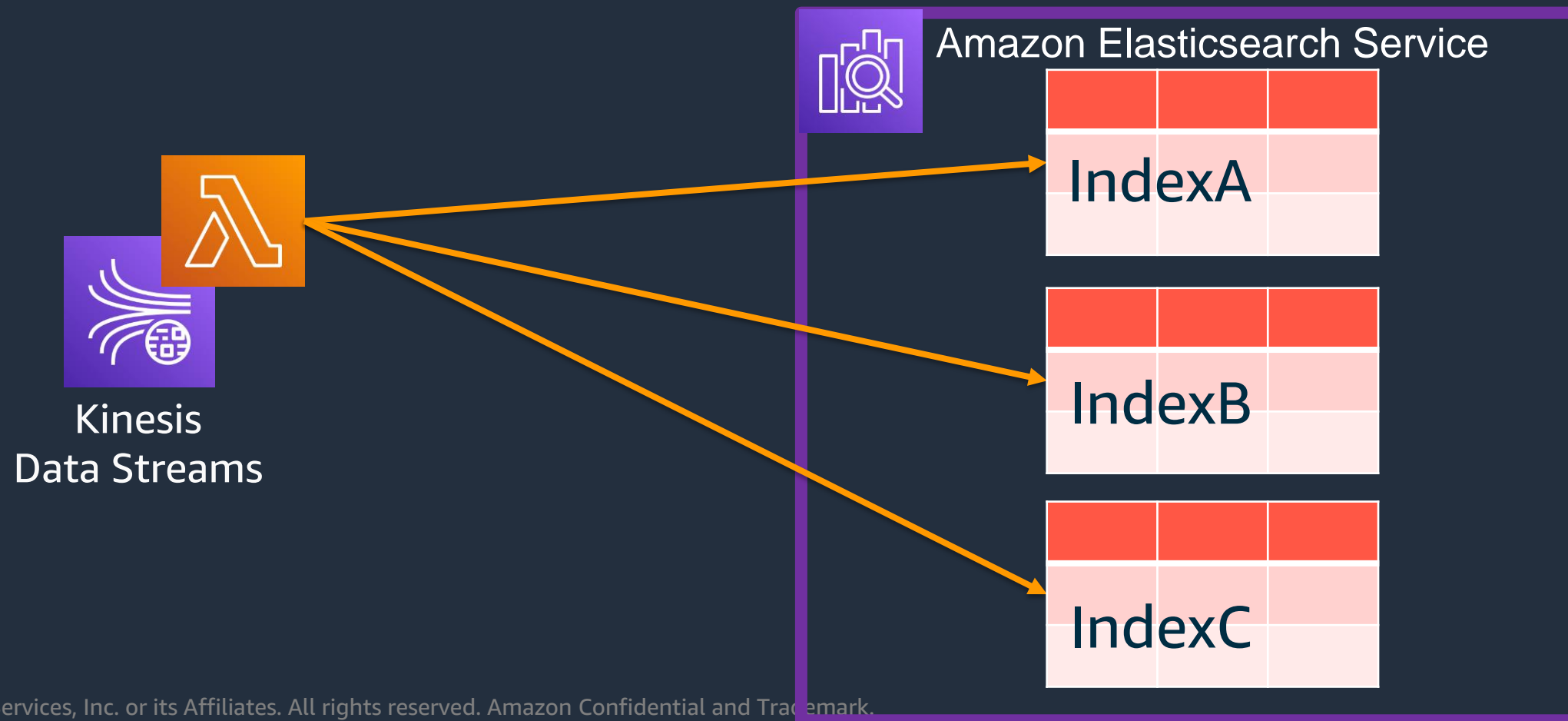


ストリームデータに対するデータ加工

- 生データから必要な項目だけを取り出して分析基盤にロードしたい
- 1分ごとにaggregateした結果を利用したい
- 他のデータソースにあるマスタデータと突合して利用したい
- 機械学習でリアルタイムに不正検知やレコメンデーションを行いたい

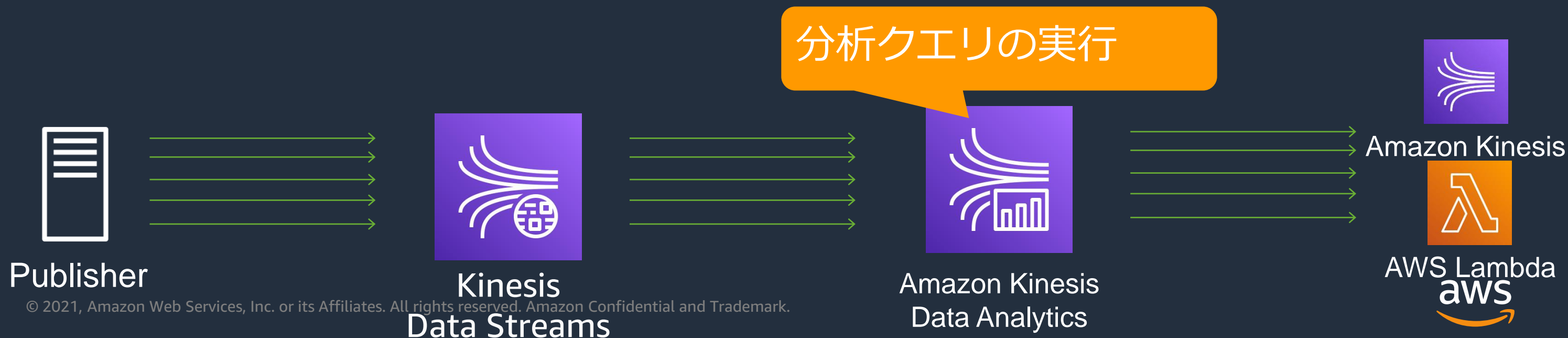
AWS Lambdaを利用したほかサービスとの連携

- Kinesis Data StreamsとKinesis Data FirehoseはAWS Lambdaを呼び出してレコードごとの加工ができる
- レコードによって別の出力先へ転送
 - イベントの種類ごとに別の集計テーブルへ格納など

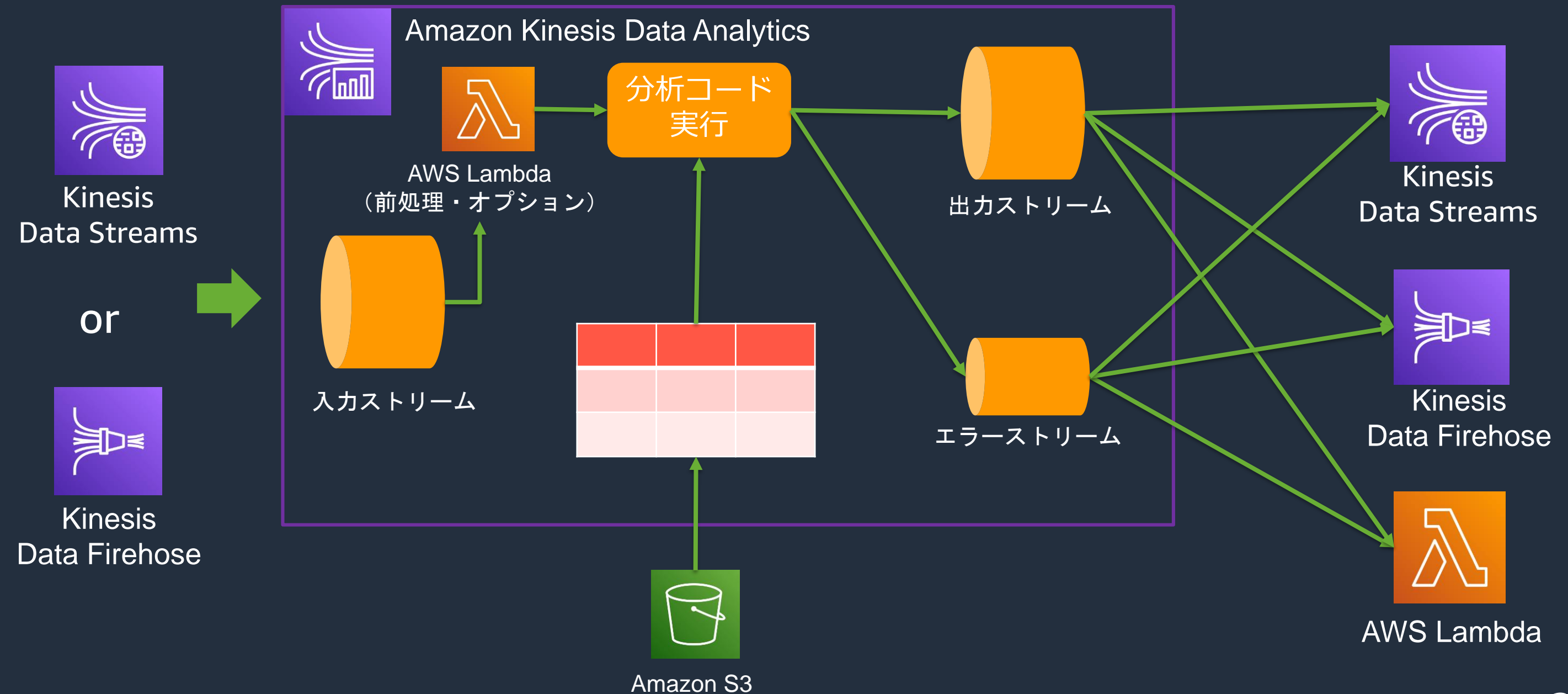


Kinesis Data Analytics

- Apache Flinkアプリケーションを構築してKinesis Data StreamsやKinesis Data Firehose内に一時永続化したデータに対して分析を行えるサービス
- 分析クエリはSQLやJava、Scala、Pythonで書ける
- 結果はKinesis Data Streams/Kinesis Data Firehose/AWS Lambdaのいずれかに出力
- 負荷に応じてKPU（コンピュートリソースの単位）を自動伸縮



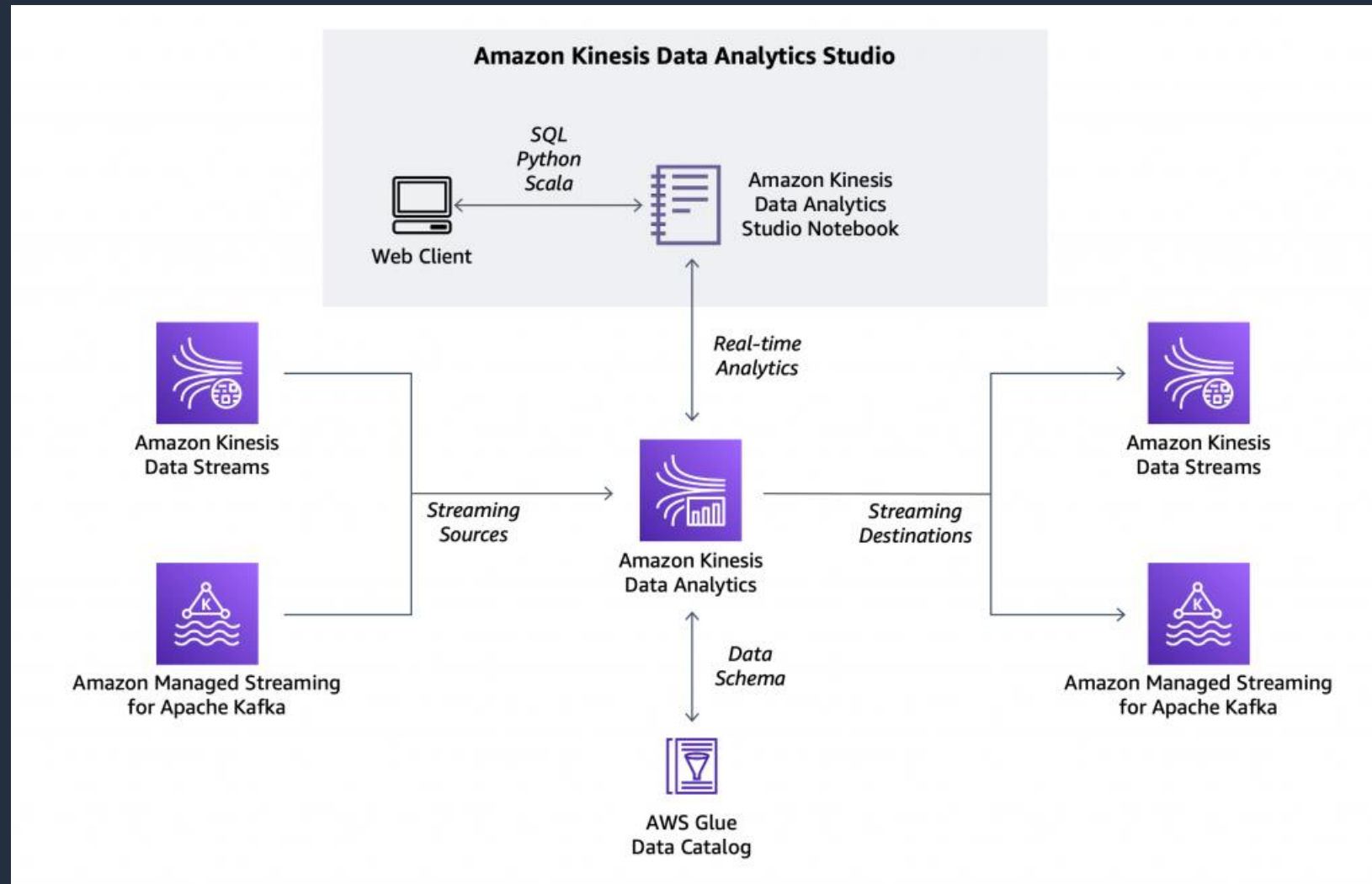
Kinesis Data Analyticsのしくみ



Kinesis Data Analytics Studio

New!

- ノートブックでストリームデータをインタラクティブに分析
- 目指した結果が得られたら数クリックで本番デプロイ



The screenshot shows the Zeppelin Notebook interface. The top bar includes the Zeppelin logo, "Notebook", a search bar, and a "Configuration" dropdown. The main area is titled "mynotebook" and contains a code editor with the following Flink SQL query:

```
%flink.ssql(type=update)
select * from stock;
```

Below the code editor, there is a table of stock prices:

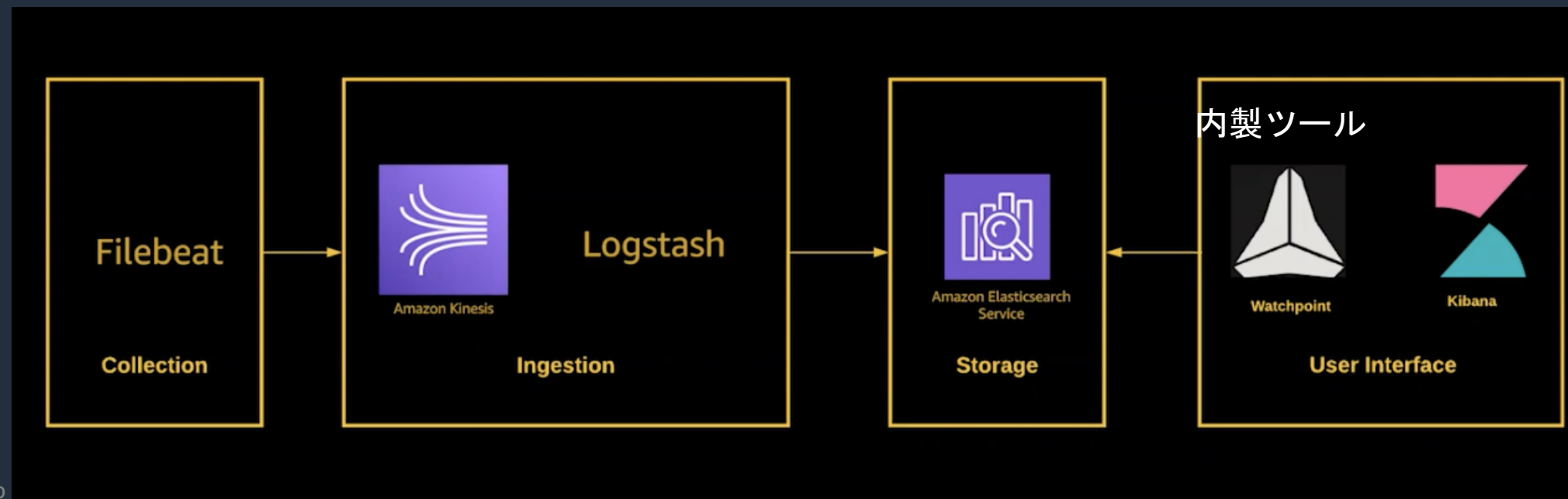
| ticker | price |
|--------|-------|
| AMZN | 98.0 |
| AMZN | 143.0 |
| AMZN | 111.0 |
| AMZN | 82.0 |
| AMZN | 46.0 |
| GOOG | 115.0 |
| GOOG | 76.0 |
| GOOG | 56.0 |

Kinesisを利用したリアルタイム 分析基盤事例

【事例】Airbnb

大規模なニアリアルタイム分析の実現

- 数百以上のサービス上で様々な方法・異なるログフォーマットで生成されるログを収集
- 1日300TB以上のデータをインデックス化
- 変動するトラフィックに対応するスケーラブルな分析環境の実現
- 複数チームにまたがるマルチテナント分析環境の実現
数秒のレイテンシで分析可能に



【事例】 THOMSONREUTERS



急激なスケールに対応する耐障害性の高いニアリアルタイム分析基盤

- 製品チームがユーザ体験を継続的に改善するための分析基盤
- Kinesis Data StreamsとKinesis Data Firehoseを利用した1つのストリームの複数用途での利用
- 1年で4,000 events/sec から10,000 events/sec へのスケールに対応
- ニュース速報時はトラフィックが通常時の2-3倍にバースト
- 新しいデータを10秒以内にダッシュボード反映
- ロンチ以来データロスやダウンタイムゼロ

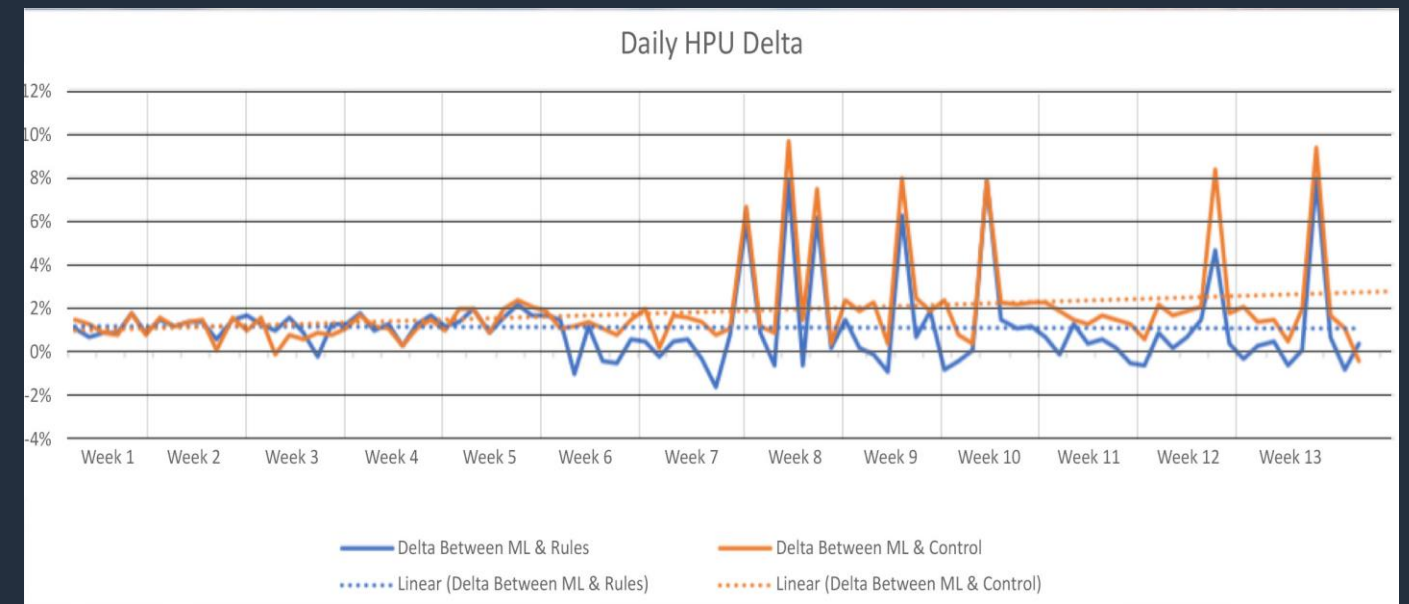
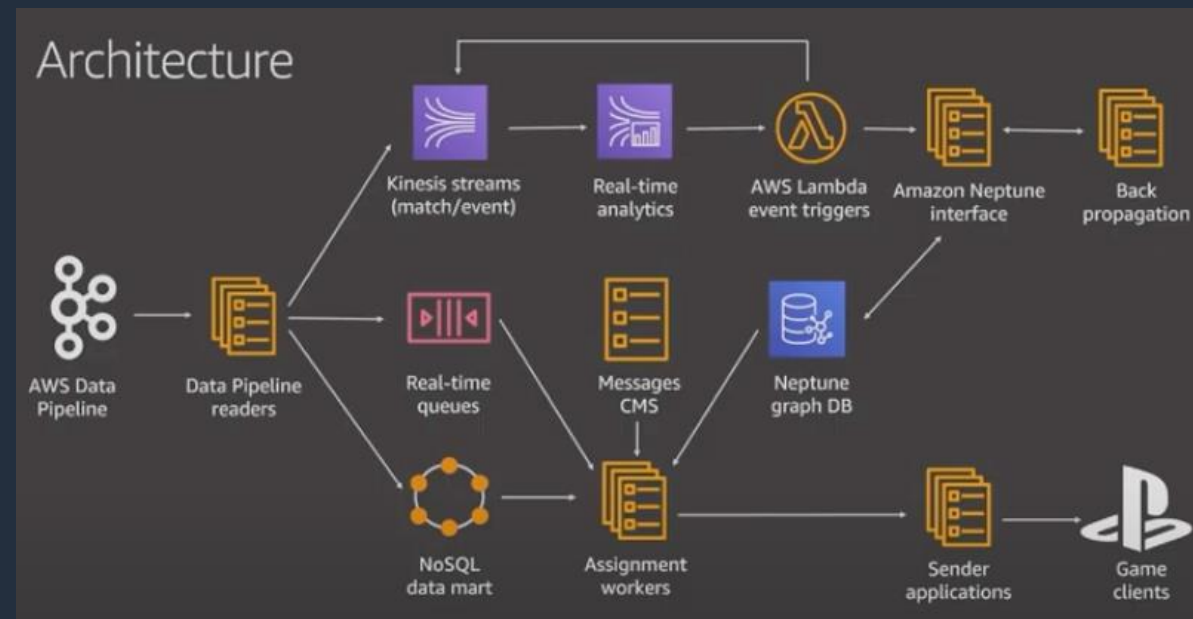


Activision

リアルタイムパーソナライズ



- FPSビデオゲーム『Call of Duty』におけるリアルタイムパーソナライズ
- Amazon Kinesisで収集したデータに対してKinesis Data AnalyticsとAWS Lambdaを用いて機械学習を用いたリアルタイムパーソナライズを行い、ユーザのプレイ時間やイベント参加率の改善を達成した



re:Invent 2019: <https://www.youtube.com/watch?v=ir33m9RJxs0>

Blog: <https://aws.amazon.com/jp/solutions/case-studies/activision-call-of-duty/>

株式会社ドリコム

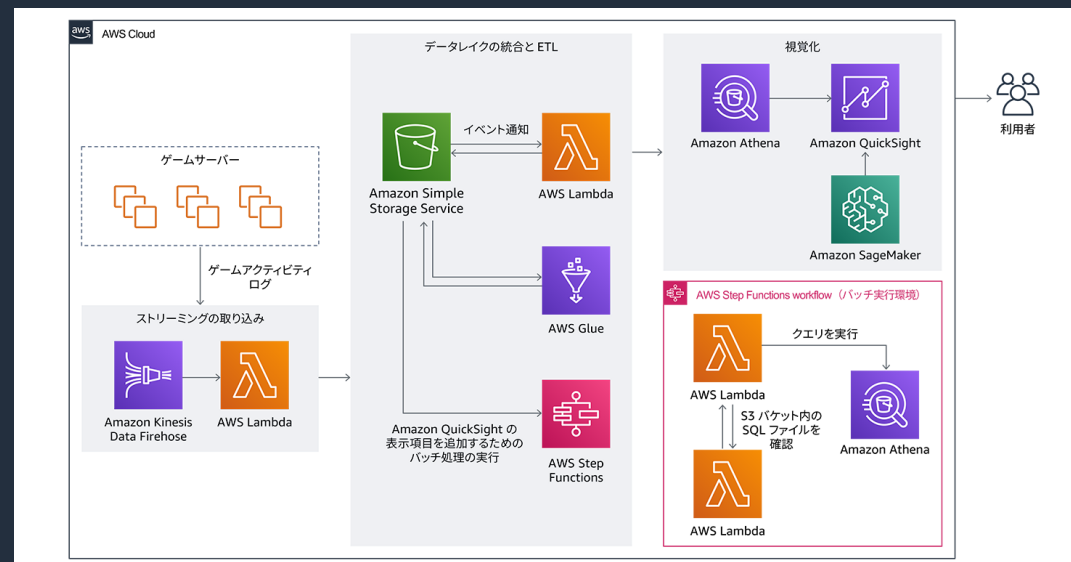
オンプレからの移行によるコスト削減



『ONE PIECE トレジャークルーズ』『みんなゴル』などのソーシャルゲームの月間250万人超のユーザ動向を捉えるデータ分析基盤を構築

チャレンジ

- Amazon Kinesis Data Firehose、Amazon Athena、Amazon Quicksightを利用して、**10分のタイムラグ**でデータ集計結果を反映
- データエンジニア**2名のみ**で複数タイトルの運用
- オンプレミス環境からの移行でシステムのコストを**60%**、人的コストを**80%**削減



【事例】 Georgia-Pacific

さまざまな製造・加工機器からのリアルタイムデータ転送



- 製紙工業の製造施設におけるさまざまな製造・加工機器からさまざまなフォーマットの構造化/非構造化データをKinesisを利用してリアルタイムにデータ収集
- アナリストはAmazon Athenaで分析
- 最初の6ヶ月で数百台の機器から約50TB（5000億以上のレコード）の転送を実現
- 機械学習による故障予測により製品を高品質に保ち、数百万ドルの利益向上を実現

まとめ

- 鮮度の高いデータを使った分析は、ビジネスを加速する
- Amazon Kinesisはクラスタ管理不要でリアルタイムにデータを利用するために必要な機能が揃ったストリーム処理サービス
- Amazon Kinesisを利用したリアルタイム分析基盤はさまざまな業界で利用されている

Thank you.

